Fig. 1	Operating System	~112
100	Computer Program Modules	~114
	Process Control Module	-116
108 110	Initialization Module	~ 118
Compostion	UID Identification Module	~ 120
102	Text Acquisition Module	122
CPU Storage Disk	Clustering Module	~ 124
	Text Modeling Module	~ 126
	Keyword Identification Module	~ 128
	Text Classification Module	~ 130
User Interface		~ 132
	Data Summarization Module	~ 134
Analysis Workstation	Data Output Module	~ 136
	Data Repository	138
	Accession Numbers	~ 140
- 	Unigene Data	~ 142
	Locus Link Data	~ 144
	Omim Web Pages	~ 146
	Literature UID Lists	~ 148
	Text Corresponding to UIDs	~ 150
	Microarray Data	~ 152
	Clusters of Accession Numbers	154
	Statistical Model of Text	156
	Key Words and Phrases	~ 158
	Text Classificati n Data	~ 160
	Summarized Data	162

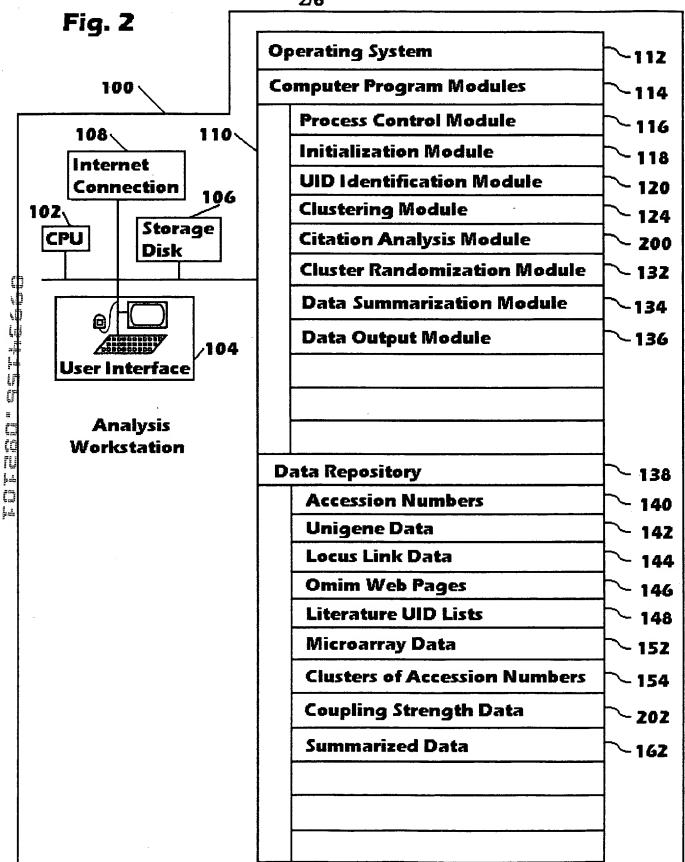
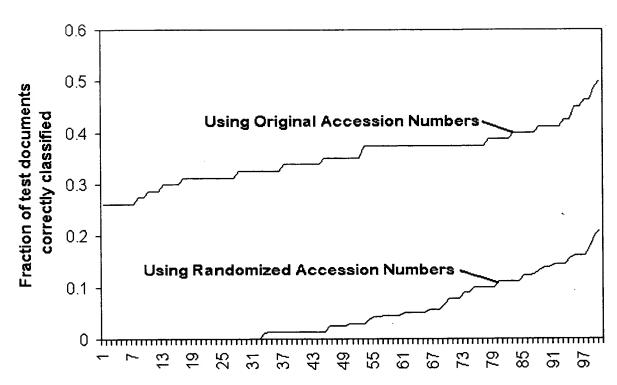
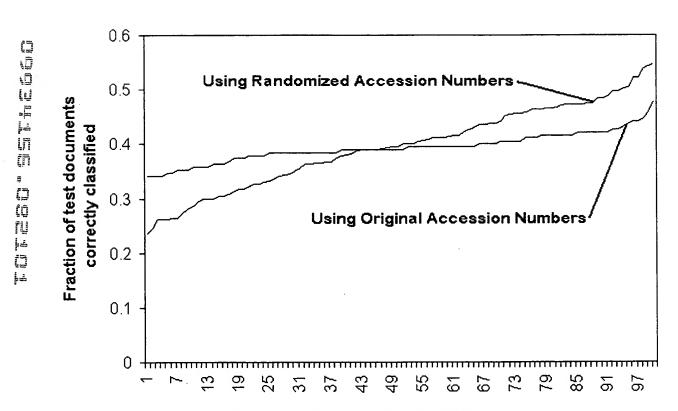


FIG. 3



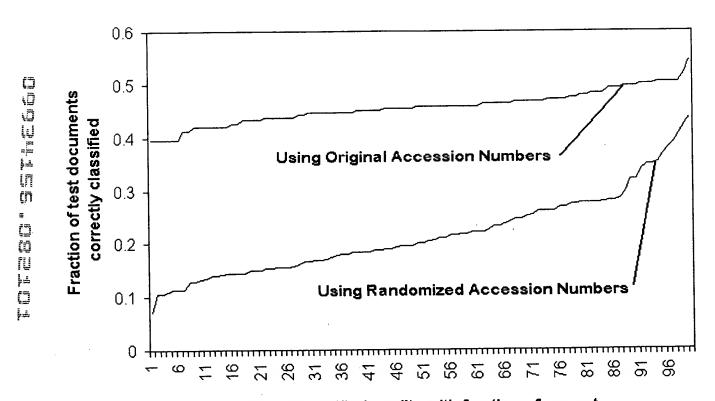
Percent test/train splits with fraction of correct classifications less than the indicated value

FIG. 4



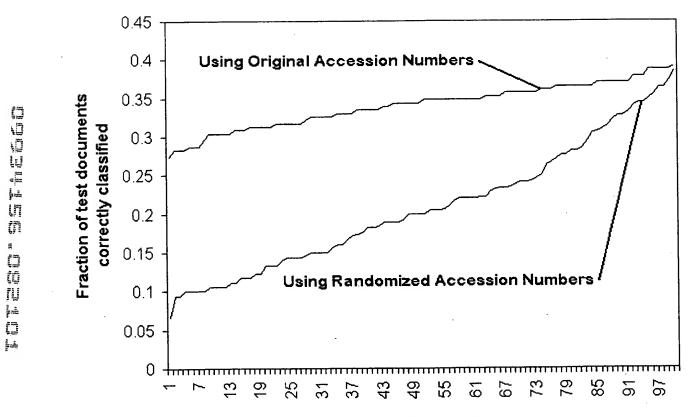
Percent test/train splits with fraction of correct classifications less than the indicated value

FIG. 5



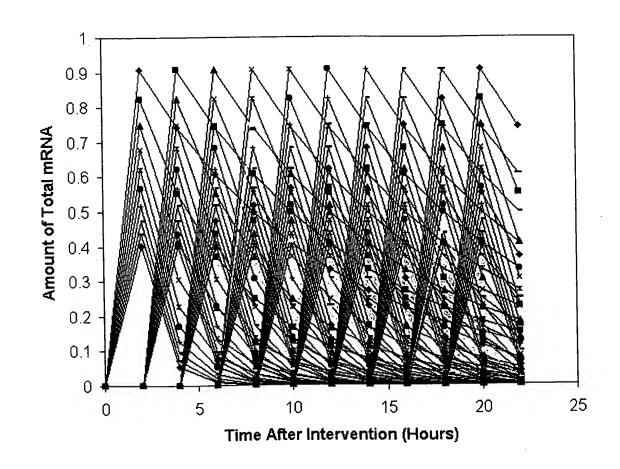
Percent test/train splits with fraction of correct classifications less than the indicated value

FIG. 6



Percent test/train splits with fraction of correct classifications less than the indicated value

FIG. 7



NGSSLACE NGBALL

FIG. 8

